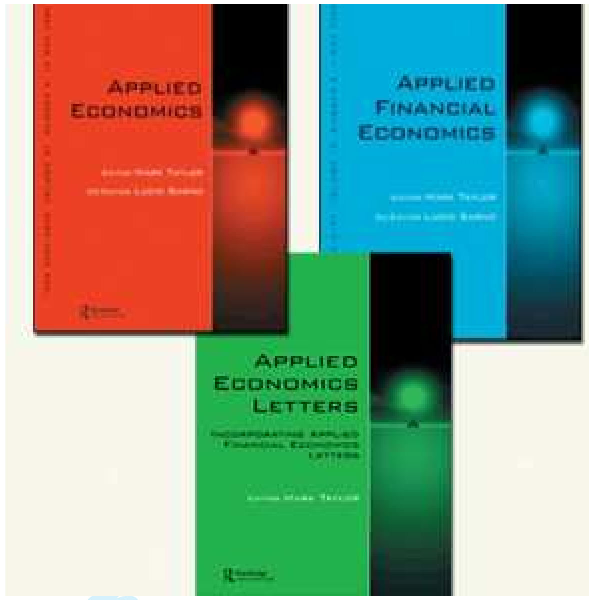
A mixed approach to payment certainty calibration in discrete choice welfare estimation

Brouwer, Roy

Postprint / Postprint
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
www.peerproject.eu

Empfohlene Zitierung / Suggested Citation:
Brouwer, R. (2009). A mixed approach to payment certainty calibration in discrete choice welfare estimation. *Applied Economics*, 1-36. https://doi.org/10.1080/00036840903035977



A mixed approach to payment certainty calibration in discrete choice welfare estimation

Journal:	*Applied Economics*
Manuscript ID:	APE-08-0425.R1
Journal Selection:	Applied Economics
Date Submitted by the Author:	05-May-2009
Complete List of Authors:	Brouwer, Roy; Institute for Environmental Studies, Vrije Universiteit Amsterdam
JEL Code:	Q51 - Valuation of Environmental Effects < Q5 - Environmental Economics < Q - Agricultural and Natural Resource Economics, D80 - General < D8 - Information and Uncertainty < D - Microeconomics
Keywords:	Payment certainty calibration, Contingent Valuation, Hypothetical bias, Econometrics

A mixed approach to payment certainty calibration in discrete choice welfare estimation

Roy Brouwer

Department of Environmental Economics, Institute for Environmental Studies, Faculty of Earth and Life Sciences, VU University Amsterdam, De Boelelaan 1087, 1081 HV, Amsterdam, the Netherlands. Phone: +31 20 5985608, E-mail: roy.brouwer@ivm.vu.nl.

Abstract

This paper provides further empirical evidence of payment uncertainty in dichotomous choice contingent valuation (CV) and proposes an alternative way of certainty calibration, moving away from conventional recoding of uncertain responses. In an international CV application, the main sources of payment uncertainty are identified related to imperfect knowledge and information about the public environmental good involved, future supply levels, income constraints, price levels and the survey instrument. Together these sources of uncertainty are responsible for a third of the error variance in the estimated discrete choice model. Accounting for the heterogeneity induced by payment uncertainty in the welfare estimation procedure with the help of a mixed probit model yields a significantly lower welfare measure albeit at the expense of estimation precision.

Running title: Payment certainty calibration in discrete choice welfare estimation

1. Introduction

Stated preference methods such as contingent valuation (CV) face a number of biases, of which payment certainty related to the method's hypothetical bias received a lot of attention after publication of the NOAA Blue Ribbon Panel recommendations (Arrow et al., 1993). Payment certainty refers to the empirical finding that respondents are unsure about their value statement for hypothetical changes in the provision level of a private or public good (e.g. Li and Mattson, 1995; Ready et al., 1995; Champ et al., 1997). Values are systematically overstated when elicited under hypothetical conditions compared to real purchase decisions, as also shown in previous work published in this journal based on private (Johannesson et al., 1998) and public goods (Veisten and Navrud, 2006). The extent to which hypothetical responses are overstated is influenced by the value elicitation method, including whether respondents were asked for their willingness to pay (WTP) or willingness to accept (WTA), and whether the good concerns a private or public good (List and Gallet, 2001).

To account for this overstatement, a number of calibration approaches have been advocated in the literature (e.g. Vossler et al., 2003). The most important ones are ex post decision ratings and the use of polychotomous elicitation formats. In the former case, the respondent is asked in a follow-up question to indicate the certainty of his WTP reply on a scale from 1 to 10 or 0 to 100 percent. In the latter case, respondents are able to express their certainty through the dichotomous choice (DC) WTP question self, for example by answering 'definitely yes, probably yes, don't know/not sure, probably, definitely no' to the presented bid amount. Typically, asymmetric approaches are applied based on self-reported payment certainty, where uncertain yes responses to a DC WTP question are recoded as certain no responses. This automatically reduces estimated mean WTP. In the limited number of studies exploring at which certainty cut-off value hypothetical WTP best simulates actual market behavior (i.e.

where uncertain yes responses are recoded to certain no responses), values vary between 6 and 10 using a scale from 1 to 10 (e.g. Champ and Bishop, 2001; Poe et al., 2002). Polychotomous elicitation formats have been used to identify similar threshold values with the help of multinomial choice models where respondents switch between certain and uncertain WTP replies (e.g. Wang, 1997; Welsh and Poe, 1998; Alberini et al., 2003). In other applications, polychotomous elicitation formats were used as a WTP follow-up question and only the 'definitely' or 'absolutely' yes responses appeared to match actual purchase behavior for private goods (e.g. Johannesson et al., 1998; Blumenschein et al., 2008). Detailed reviews of the effect of payment certainty calibration approaches on stated WTP are presented in Samnaliev et al. (2006), Chang et al. (2007) and Shaikh et al. (2007).

Although there exists no consensus in the literature about the most appropriate payment certainty elicitation format, the available empirical evidence listed above suggests that both approaches can help to reduce hypothetical bias in stated preference research. However, the evidence is limited and a fair share of the studies focus on private market goods. More research is needed in the area of public good valuation, also regarding underlying sources of payment uncertainty, which are rarely investigated (Alberini et al., 2003), but expected to provide important insight into the design of more reliable stated preference survey formats and WTP values. This paper further investigates the issue of payment certainty in an international CV application to value a public environmental good. The study's main objective is to test the effect of an alternative certainty calibration method, moving away from existing recoding procedures and taking full account of the self-reported measurement error caused by payment uncertainty in the stochastic error component of a mixed probit discrete choice model. The purpose of this study is not to address appropriate payment certainty measurement methods, but rather assess what drives and explains payment certainty, use the available information in an integrated way in the choice model's error variance to calibrate stated preferences, and compare the results with the conventional calibration approach.

The remainder of this paper is organized as follows. Section 2 presents an overview of the existing literature on explaining payment certainty. This is followed in section 3 by a description of the case study presented in this paper. The WTP results and self-reported payment certainty are presented in section 4, the estimated regression models in section 5 and the results of alternative certainty calibration methods on welfare estimation in section 6. Underlying reasons for the observed variation in stated payment certainty and the results of the estimated econometric model explaining payment certainty are discussed in section 7. Section 8 concludes.

2. Explaining payment certainty

The empirical evidence in the CV literature over the past fifteen years suggests that respondents are uncertain about their stated preferences for changes in the level of public environmental good provision. Although there exists no single unifying theoretical model that explains why people know their preferences with certainty (or not), familiarity and experience with (public environmental) goods and their valuation are generally assumed to be important determinants of preference stability and certainty. Hoeffler and Ariely (1999) show that preference stability is positively correlated with choice experience and choice effort (easy versus hard choice). In the latter case, a higher level of effort leads to more stable preferences, but less preference strength, meaning that respondents facing a hard choice are less certain of their preferences than respondents facing an easy choice. Through repetition respondents are expected to be capable of making more precise and consistent decisions, because they learn about the survey format, associated (hypothetical) market environment and their preferences

(List 2003)[1]. Ajzen's theory of planned behavior (Ajzen et al., 2004) has also been suggested as an important starting point to better understand payment certainty. The stronger the behavioral intent (measured through greater certainty underlying stated WTP), the stronger the link to actual behavior (Blumenschein et al., 2008).

The important role of a respondent's prior knowledge of the good in question to reduce payment uncertainty is evidenced in Loomis and Ekstrand (1998). Respondent familiarity with the environmental good is the only significant explanatory factor together with bid price. In a mail survey, US households were asked for a voluntary contribution in terms of WTP for preserving the Mexican Spotted Owl and its habitat. Following a DC WTP question, a post-decision certainty scale from 1 (not certain) to 10 (very certain) was used to elicit the level of response uncertainty. An OLS regression model on the pooled data (both yes and no responses) was applied to determine the sources of variations in the self-reported certainty scores. The other (non-significant) explanatory factor included in the model was a dummy variable indicating whether the respondent had ever visited the area proposed for protection. A significant quadratic effect of bid price was found, implying that self-reported payment certainty is highest around the lowest and highest bids, and lowest for intermediate bid levels closest to expected maximum WTP.

Champ and Bishop (2001) examined US household preferences for a voluntary wind energy program from a local private electricity provider in another mail survey in Madison, Wisconsin. A split sample approach was used to identify possible differences in household behavior under a hypothetical and actual payment scenario. The same scale was applied as in

[1] In turn, the more uncertainty in someone's preferences, the more expressed preferences will be subject to procedural and descriptive influences (e.g. Schkade and Payne 1994; Ariely et al., 2003).

Loomis and Ekstrand (1998) to measure the level of certainty for the group of respondents who were asked to pay for wind generated electricity under the hypothetical scenario. In an ordered probit model, respondent attitude to the proposed program was responsible for a large part of the observed variation in the self-reported certainty scores. Respondents in favour of the program and willing to pay the extra cost expressed higher certainty levels. No other significant explanatory factors were detected.

Similar results were found in a mail survey in New Hampshire and Idaho where US households were asked to pay a user fee to access public land (Samnaliev et al., 2005). An ex post rating scale and polychotomous certainty choice format were used in two different samples. Responses to the former follow-up certainty question were regressed on possible explanatory factors. Two separate logistic regression models for the yes and no responses were estimated, where the dependent variable took the value 1 if the certainty score equalled 10 (very certain). Respondents who objected against the imposed user fees (usually referred to as protest response in CV) were more certain in rejecting the bid price than others, reflecting (as argued by the authors) respondent general attitude towards the hypothetical market and the environmental protection program being valued.

Finally, empirical evidence supporting the relationship between respondent attitude towards a public environmental good and payment certainty levels is also found in Akter et al. (2009), who asked international air travelers at Schiphol airport in Amsterdam, the Netherlands to pay a carbon travel tax to offset carbon emissions from flying. When changing the market compliance imperative from a mandatory carbon tax to a voluntary contribution, a third of all air travelers considered it unlikely they would actually pay the price they said they would pay. In an ordered probit model, bid price, respondent sense of responsibility and belief in the effectiveness of the voluntary carbon market were found to be the main determinants of self-reported payment certainty. The higher the bid price, the less likely someone would actually pay, while a respondent's sense of responsibility for contributing to climate change and respondent belief in the proposed tree plantation program to mitigate climate change resulted in a higher likelihood of paying.

3. Case study design

The data used in this paper are taken from an international water quality CV study conducted in the Scheldt river basin. The Scheldt is 350 km long and flows through three countries. The river originates in France, runs through Belgium and ends in the Netherlands where it flows into the North Sea. The international river basin covers an area of over 36 thousand square kilometres and has almost 13 million inhabitants. This corresponds with an average population density of 350 inhabitants per square km, which is almost three times the European average. The Scheldt is one of the pilot river basins in the implementation of the European Water Framework Directive (WFD) adopted in 2000, aiming to improve water quality of all European freshwaters to 'good ecological status' by 2015. Current Scheldt water quality does not meet this objective and was classified as 'moderate to poor' in the first WFD regulatory assessment report (Scaldit, 2006). In order to elicit public preferences for water quality improvement and obtain measures of WTP for the WFD water quality objectives, a CV questionnaire was sent out in October 2005 to a random selection of 17,000 households across the Scheldt basin. In particular, questionnaires were sent to 5,000 households in Artois-Picardie in northern France, 9,000 in Flandres, Belgium, and 3,000 households in the Dutch part of the basin.

The common survey design was developed together with water experts from the responsible water management authorities and was pre-tested in French, Flemish and Dutch. The questions are identical in the three versions of the questionnaire except for the description of the current situation, which was modified to the specific prevailing circumstances and conditions in the three different parts of the river basin. A map and common water quality ladder were used to depict the current situation and show respondents the location and quality levels of the river relative to their place of residence.

The questionnaire consists of four parts. In the first part, respondents are asked about their water related recreation activities, followed by a series of questions about their perception of current water quality and the importance they attach to water quality. Following these introductory questions, respondents are presented with a one-page information statement in the second part of the questionnaire, in which the actual water quality situation is described with the help of a map and a brief explanation of the WFD. After the information statement, respondents are asked how familiar they are with the presented information and how important it is to them that the WFD objective of 'good ecological status' is reached. This is then followed by a DC WTP question using 10 different bid amounts ranging between 5 and 250 Euros and a post-decisional payment certainty question on a scale from 0-100% with 10 percent intervals. It is explained that 0 means not certain at all and 100 means completely certain. An open-ended follow-up question is used to enable those respondents who are not completely certain to specify why not. The WTP question and the follow-up payment certainty questions are reproduced in Annex 1 of this paper. The third part of the questionnaire consists of a series of questions about the respondent's demographic and socio-economic household characteristics to test the survey's representativeness. The fourth and final part concludes with a number of questions about the questionnaire self and in particular the WTP question to examine respondent understanding of the WTP question, the perceived credibility of the valuation scenario and the difficulty experienced in answering the WTP question.

The overall response rate was 18 percent, which is not unusual for these kinds of surveys in the three countries where the survey was conducted. As expected (based on available country specific demographic and socio-economic statistics), respondents differ significantly between countries except for age. Although the response rate was not very high, the three samples approximately represent the average inhabitant of the three countries.

4. Hypothetical WTP and payment certainty

Respondents were asked for their WTP through extra annual taxes until and including 2015 to reach a good ecological status for all water bodies in their part of the Scheldt basin. Just over half of all respondents (52%) said they were willing to pay extra for this purpose, and this share is more or less the same across the three river basin countries. Most of the respondents who were not willing to pay extra for the proposed water quality improvements motivated this by saying they lacked sufficient income (15%), 'the polluter should pay' (7%), 'I already pay (enough) taxes for water quality' (7%), and 'there are more important other things I prefer to spend my money on' (5%). About five percent believe that current water quality is good enough or consider water quality improvements not important enough to pay for. The share of respondents protesting against the WTP question in this study based on considerations such as the polluter should pay, lack of trust in the feasibility of the proposed program of measures or the responsible authorities is 9.5 percent of the total sample[2]. Protest rates vary slightly across

[2] Protest bidders typically object against the imposed market structure in a CV study (e.g. Meyerhoff and Liebe, 2006). A separate analysis was carried out using the non-parametric Mann-Whitney test to test differences

the three countries, from 7 percent in Belgium to 11 percent in the Netherlands and 13 percent in France.

The probability distribution function of the positive responses to the DC WTP question is presented in Figure 1. As expected, the higher the bid amount, the lower the probability that the respondent is willing to pay. The certainty experienced when answering the DC WTP question is also presented in Figure 1. A distinction is made between yes and no votes. A small, but significant negative correlation exists between bid price and self-reported certainty for the yes votes and a significant positive correlation for the no votes. Corresponding with the results found by Chang et al. (2007), yes voters are more confident overall about their answer than no voters. Self-reported payment certainty surrounding the yes replies decreases as the bid price increases from 5 to 100 Euros, rises somewhat at 150 and 200 Euros and then drops again at 250 Euros.

INSERT FIGURE 1 HERE

Respondents are significantly more certain that they are willing to pay the lowest bid (€5) than the highest bid (€250) (Mann-Whitney Z=-4.886; $p<0.001$). No significant difference can be found between the payment certainty levels for respondents not wanting to pay the lowest and highest bid. The variation in payment certainty around the no responses is also less compared to the yes replies, ranging between 50 and 63 percent around a mean of 55 percent. Finally, payment certainty is significantly lower in France (average certainty is 71% for yes

---

between protest and non-protest bidders. Protest respondents are significantly higher educated and wealthier male respondents from slightly bigger households, with less trust and confidence in the description of the current situation and the feasibility of reaching the WFD objective of good ecological status for all water bodies in their region (test results are available from the author).

votes) compared to Belgium and the Netherlands (average certainty of yes votes is respectively 84 and 82%). No significant differences can be found between Belgium and the Netherlands (Mann-Whitney Z=-0.217; p<0.828 for the pooled yes and no responses).

5. Estimated regression models

Standard probit regression models were used (in Stata 10) to estimate mean WTP values (e.g. Cameron and James, 1987) and evaluate how different certainty calibration approaches affect these values. Two different types of models are estimated: a fixed effects probit model where self-reported payment certainty is included as a common explanatory factor and a random effects probit model where payment certainty is included as a separate stochastic error source to account for preference heterogeneity[3]. In the first model it is assumed that payment certainty affects WTP choices across all individuals in the same way, whereas in the second model the underlying assumption is that a significant share of preference heterogeneity is driven by groups of respondents with different WTP values separated from each other by payment certainty. Testing for random effects has the advantage that it explicitly shows the model variance associated with payment certainty. This can then be compared with the results from recoding WTP values based on different certainty cut-off points in the fixed effects probit model. The different factors that appeared to have a significant impact on stated WTP are presented in Table 1 (the calculated welfare estimates are presented in Table 2 in the next section).

[3] The model used here is a mixed binary probit model where random preference heterogeneity is picked up in the structure of the covariance matrix (Train, 2003) due to the expected clustering of responses across self-reported payment certainty levels (not to be confused with the panel data structure found in repeated choice experiments). A short model description is presented in Annex 2.

INSERT TABLE 1 HERE

The estimated probit models in Table 1 are highly significant as shown by the outcome of the Wald tests. Individual coefficients are significant at the one percent level unless indicated otherwise. Two different fixed effects probit models are presented. The first model includes the original self-reported certainty levels as an explanatory factor and shows that certainty significantly affects stated WTP in a positive way: the higher (lower) respondent certainty, the higher (lower) the likelihood of agreeing to pay the bid price. As expected, bid price has a significant negative effect on stated WTP (the higher the price, the less likely someone is willing to pay). An interaction term is included between payment certainty and bid price to test if an effect exists of payment certainty variation on respondents' WTP higher or lower bid prices[4]. The estimated coefficient indicates that there exists an additional negative price effect: respondents who are more certain are less likely to agree to pay a higher bid price than respondents who are less certain.

In the second probit model, the variable 'certainty' is included as a dummy to test whether respondents who are completely certain are more willing to pay than respondents who are not completely certain. The dummy variable has the value one if respondents are 100% certain, so respondents who are completely certain are, ceteris paribus, more likely to pay than respondents who are not 100% certain. An interaction term is again included between the certainty dummy and bid price to test whether respondents who are completely certain also have a significantly different WTP value than respondents who are not completely certain. As in the first model, the interaction term has a significant downwards effect on mean WTP.

---

[4] In order to be able to compare the impact of the interaction term with that of the bid price, the original payment certainty classes were in this case recoded between 0 and 1.

The third model is the random effects probit model where the original payment certainty classes are included as a random variable to take into account possible variance due to the clustering of responses around stated payment certainty. This variance due to self-reported certainty is significant as reflected by the significance of the standard deviation of the random variable. Respondents answering the DC WTP question can be classified into more or less homogenous groups based on the certainty intervals. The correlation coefficient measuring the fraction of the total variance of the error terms explained by payment certainty (33%) is statistically significant based on the outcome of the Likelihood Ratio test (chi-square=173.86; $p<0.001$). The Ben-Akiva and Swait (1986) test for non-nested choice models is used to test the statistical significance of the improvement of model fit of the random effects model over the fixed effects model. The probability that the goodness of fit measure of the latter outperforms that of the former is virtually zero ($p \leq \Phi(-171.187) \approx 0$ with Φ being the standard normal cumulative distribution function). These results hence show that stated payment certainty is responsible for a significant share of model error variance, and accounting for this error variance significantly improves the model fit.

The other significant explanatory factors included in the regression models are a combination of theoretically expected and empirically driven factors. The former display the expected direction of influence and the coefficient estimates are more or less constant across the three models. The theoretically expected factors include household disposable income (the higher disposable income, the higher the probability of WTP), respondent attitude towards water quality improvements (measured through the importance attached to reaching good ecological water status; respondents who believe reaching good status is very important are more likely to pay), familiarity with the information provided in the questionnaire about current water quality (respondents who are more familiar with the information provided are more likely to pay)[5], and respondent belief in the valuation scenario (the more someone believes the scenario, the higher the likelihood of WTP)[6].

Ad hoc variables include the country dummy variables for Belgium and the Netherlands, which show that the estimated WTP functions are not the same (and hence not transferable) across the three countries making up the international river basin. The difference between the parameter estimates for Belgium and the Netherlands is not significant in any of the three models (based on the Wald test; test results are available from the author). This indicates that French river basin residents value the benefits from water quality improvements significantly different from residents in Belgium and the Netherlands (the latter are less likely to pay). Respondent age is the only significant demographic factor. No a priori expectations existed regarding its direction of influence. In this case, older respondents are more likely to say yes to the presented bid amount than younger respondents. Other demographic and socio-economic characteristics have no significant impact on stated WTP.

6. Accounting for payment certainty in welfare estimation

Welfare estimates are presented in Table 2, including the uncorrected mean WTP based on the first fixed effects probit model in Table 1, mean WTP based on different certainty calibration cut-off points in the first fixed effects probit model, mean WTP for those respondents who are and those who are not 100% certain based on the second fixed effects probit model in Table 1,

[5] The dummy has the value 1 if respondents never heard of the information before.

[6] This variable was measured on a 5-point semantic differential scale ranging from 'completely not credible' to 'completely credible'. Dummy variables were included for the two positive levels (somewhat and completely credible). No significant difference can be detected between the two dummy variables in any of the three models based on the Wald test (test results are available from the author).

and mean WTP derived from the random effects probit model presented in Table 1. The 95 percent confidence intervals around estimated mean WTP are based on the delta method (Greene, 2003). A simple two-tailed *t*-test is used to test the statistical significance of the observed differences between the welfare estimates. The statistical efficiency of the welfare estimates is measured with the help of the mean squared error (MSE).

INSERT TABLE 2 HERE

A first important observation from Table 2 is that respondents who are 100% certain about their stated WTP are willing to pay, on average, significantly more than respondents who are not 100% certain (t = 3.888; p<0.001). The negative additional price effect found in the second probit model for respondents who are completely certain (see section 5) is offset by the fact that these respondents are more likely to pay overall than respondents who are not completely certain. Not accounting for payment uncertainty hence results in an overestimation of the welfare measure. The statistical efficiency of the welfare estimates is also highest for respondents who are 100% certain when examining the MSE values[7]. This implies that WTP values are most accurate when respondents are completely certain.

A second observation is that, as the MSE show, the statistical inefficiency of the calibrated welfare estimates (i.e. recoding of uncertain WTP based on different certainty cut-off points) increases as the restrictions imposed on payment certainty are more stringent (and the welfare estimates decrease accordingly as expected). Hence, as demand for payment certainty of hypothetical WTP responses increases for welfare estimation purposes, the practice of recoding uncertain responses results in gradually less precise welfare estimates.

[7] Despite the lower number of observations underlying this estimate.

A third observation is that accounting for the different constant term across respondents with different payment certainty levels in the random effects model yields a significantly lower welfare measure compared to the uncorrected welfare estimation procedure ($t = 2.440$; $p<0.015$). In the latter case, the assumption is that the constant is the same across all respondents irrespective of respondent certainty. The more conservative welfare measure comes at the expense of estimation efficiency. In this particular case, also the variation coefficient of both welfare estimates were compared to double check the big difference in absolute MSE values[8]. Based on this alternative statistical efficiency measure, the difference is less dramatic. The variation coefficient of the welfare estimate derived from the random effect probit model (7.5%) is about 60 percent higher than the variation coefficient for the uncorrected welfare measure (4.6%), but still relatively low.

Finally, comparing the lower random effect welfare estimate with the calibrated welfare estimates, it compares best with the estimate obtained when responses have to be at least 50% certain and are recoded otherwise ($t = 0.653$; $p<0.515$). The random effect welfare estimate is significantly higher than the welfare estimates calibrated upon a cut-off point of 60, 70 and 80 percent. A remarkable finding is that the random effect model generates a welfare estimate, which is not very much different from the welfare estimate based on the uncertain (<100%) responses only ($t = 0.124$; $p<0.901$). An intuitive explanation for this finding is hard to give. The random effect model includes respondents who are completely certain and respondents who are not. The weighted average of both groups (respondents who are 100% certain and respondents who are not) is €103.6, which is (as expected) closer to the uncorrected mean WTP.

[8] The variation coefficient is equal to the standard error divided by mean WTP.

7. Underlying sources of payment uncertainty

Respondents who were not 100% certain about their stated WTP were asked in an open-ended follow-up question why not. The reasons given were carefully analyzed and categorized, followed by the estimation of a formal econometric payment certainty model like the ones discussed in section 2. The main sources of uncertainty underlying stated WTP are presented in Table 3. Payment uncertainty can be related to imperfect knowledge and information about (i) the good to be valued (which is a function of information provision and experience), including its provision level now and in the future (referred to here as 'supply uncertainty'), (ii) the utility derived from different 'consumption' levels (referred to here as 'demand uncertainty'), which is a function of individual respondent characteristics such as household income levels and corresponding purchasing power (now and in the future as CV research often asks respondents to pay over a specified period of time in the future), (iii) particular simulated market conditions (referred to here as 'survey instrument uncertainty'), such as respondent trust in property right security when paying for a public good, public good suppliers (e.g. government or other) and related payment mechanisms (e.g. tax or user fee), and (iv) price levels.

INSERT TABLE 3 HERE

Almost half of all self-reported payment uncertainty (45%) is related to the survey instrument, followed by respondent uncertainty about his or her future income situation (mentioned by a fifth of the uncertain sample), and current and future price levels (17%). A high cost price can be seen as a choke price and hence as another demand related source of uncertainty. Within the category 'survey instrument uncertainty', three sub-groups are distinguished. Policy scenario uncertainty constitutes the largest source of uncertainty. This includes lack of trust that the money paid will actually be spent on the improvement of water quality, lack of trust in the government and water managing institutions as the main providers of the good for which respondents were asked to pay, the perceived inefficiency of public administration and lack of control over how public money is spent. This category is closely related to and overlaps with the 'supply' related sources of uncertainty as it also refers to doubts about the provision of the environmental good in question. Another source of survey instrument uncertainty relates to the appropriateness of taxes as the payment mode or whether individual households are the right target group for this particular problem.

Although it was emphasized in the questionnaire that the contingent market simulation is based on the polluter pays principle, fifteen percent of the uncertain respondents question whether this is actually the case, including how much surrounding countries will do to solve the problem and to what extent all households will pay for this. Together with the lack of trust in the government, these latter reasons are usually classified as protest response in CV. Protest beliefs may hence be an important source of payment uncertainty. On the other hand, respondents who are uncertain may also be more inclined to resort to protest beliefs when trying to explain why they are uncertain due to instable preferences.

Interesting differences were found when looking at the main sources of uncertainty across the three countries. In France, doubts about household income are the main source of uncertainty, and also in the Dutch sample most respondents are worried about their future employment status. These concerns are much less a source of uncertainty in Belgium where most respondents lack trust in the authorities and doubt that the money will actually be spent on the improvement of water quality. This reason comes second in the Netherlands, and plays almost no role in France. Also the feasibility of reaching a good ecologic water status plays an important role in Belgium and the Netherlands, but much less in France. Uncertainty about whether the polluter will pay plays a stronger role in France than in Belgium and the Netherlands. Related to this, a remarkable finding is that uncertainty about what surrounding countries will do is mentioned by almost 10 percent of all respondents in the Dutch (downstream) sample, less than one percent in Belgium and never in the French (upstream) sample. The high bid price is an important source of uncertainty in all three samples.

Based on these identified sources of uncertainty, the observed variation in self-reported payment certainty was regressed on a combination of related explanatory factors in an ordered probit model[9]. The statistically best fit results for the pooled (yes and no) WTP responses are presented in Table 4 (only explanatory factors that were found to be significant at the ten percent level are included). Coefficient estimates are significant at the one percent level unless indicated otherwise. Besides the pooled model results, the results for the positive and negative WTP responses are also presented in Table 4. Including a dummy for the positive and negative DC WTP responses in the pooled model has a highly significant positive impact on self-reported payment certainty (i.e. respondents who were willing to pay were significantly more certain), but this variable is correlated with, for instance, the bid price and household income. To avoid multicolinearity, the results are presented in two separately estimated models.

INSERT TABLE 4 HERE

[9] The payment certainty measurement scale is interpreted as an ordinal scale, justifying the use of an ordered probit model. This means that 100 percent is more certain than 50 percent, but 100 percent is not necessarily twice as certain as 50 percent. Nor is an increase in certainty from 50 to 60 percent the same as an increase in certainty from 90 to 100 percent. There are simply 11 categories with 10 percent being more certain than 0 percent, 20 percent being more certain than 10 percent etc.

In order to test the sensitivity of the estimated model results to the number of cut-off points, different classifications were used. The pooled model presented in Table 4 yields the best model fit and is based on recoding of the original payment certainty categories in 4 classes: less than 50% certain, 50% certain, between 50 and 90% certain and 100% certain[10]. Contrary to respondents who were willing to pay the presented bid amount (yes voters), certainty levels for respondents who were not willing to pay (no voters) are not in any way affected by the supply related indicator (confidence in environmental good provision) and the level of the bid price. Respondent confidence in the feasibility of environmental good provision increases payment certainty underlying stated WTP except for those respondents who refused to pay. The non-linear price effect also applies only to the positive WTP responses. In general, the higher the bid amount respondents are asked to pay, the less certain they are about their WTP. The small, but statistically significant positive quadratic effect implies a U-shaped curve when plotting the predicted payment certainty against the bid price (e.g. Loomis and Ekstrand, 1998). That is, certainty is highest around the lowest and highest bids and lowest for intermediate bid levels where mean WTP is expected to be.

As expected, knowledge has a significant impact on payment certainty irrespective of a respondent's WTP reply (respondents who are not familiar with the information in the questionnaire are less certain about their WTP responses). The same applies to ability to pay as an important driving force behind WTP (the higher the income level, the more certain the respondent is about his WTP), and protest against the survey instrument (protesters are more

[10] The same explanatory factors are found to be statistically significant based on the original certainty categories and OLS regression based on these original categories, indicating that the estimated model is robust. The explanatory power of the estimated OLS models (adjusted R-square) is 35% for the yes responses and 37% for the no responses.

certain about their stated WTP)[11]. Whereas the effect size of knowledge is the same for both yes and no responses, it is two to three times higher for yes than no responses in the case of income and protest.

Finally, other significant factors include respondent gender, age and country of residence. Women are less certain than men (based on yes responses only), and older respondents are less certain than younger respondents (irrespective of their WTP reply). Respondents from Belgium and the Netherlands are significantly more certain about their stated WTP than respondents from France, whilst accounting for a variety of other influencing factors. The differences between the parameter estimates for Belgium and the Netherlands are statistically significant except for the no responses, suggesting that social-cultural differences may have played a role between the samples[12].

8. Conclusions

This paper's main objective was to present the effect of an alternative payment certainty calibration method on welfare estimation based on hypothetical WTP, moving away from existing recoding procedures and taking full account of the self-reported measurement error caused by payment uncertainty in the stochastic error component of a mixed probit discrete choice model. The study confirms what other CV studies found before, that is, that respondents face considerable uncertainty when participating in a simulated hypothetical market, and this uncertainty significantly affects their stated WTP. Not accounting for

[11] Protest refers here to the underlying reason for payment certainty, not to the WTP question, hence the reason why the effect is significant for both yes and no responses.

[12] The Wald chi-square equals 3.21 (p<0.073) for the pooled model, 5.21 (p<0.022) for the yes responses and 0.08 (p<0.776) for the no responses.

payment uncertainty results in an overestimation of the welfare measure. Comparing respondents who are and respondents who are not completely certain about their stated WTP, the latter are significantly less willing to pay than the former and the estimated welfare measure is less precise. Preference heterogeneity in stated WTP due to payment certainty is significant and accounting for this heterogeneity in the mixed probit discrete choice model yields a significantly lower WTP estimate, comparable to the calibrated welfare estimate when responses have to be at least 50% certain. An important advantage of the mixed model over the recoding of uncertain responses in a common effect probit model is that it explicitly shows the model variance associated with payment certainty. Furthermore, accounting for the error variance caused by different levels of payment certainty in the mixed probit model significantly improved the model fit.

Besides the application of an alternative certainty calibration model, also the mix of a structured qualitative analysis and econometric modeling of underlying sources of payment uncertainty is new in this study. Sources of payment uncertainty were related to imperfect knowledge and information about the environmental good involved, future supply levels, income constraints, price levels and the survey instrument. Together these sources of uncertainty were responsible for a third of the error variance in the estimated discrete choice model. The strong correlation between payment certainty and respondent familiarity with the public environmental good in question and belief in the presented valuation scenario points out the importance of the role of information in stated preference research.

In conclusion, the results presented in this paper indicate that payment certainty is a significant random error component, but not a random process in itself. In-depth examination of the underlying sources of payment uncertainty provides important signals as to how to improve stated preference survey formats and produce more reliable WTP values. Taking full advantage of the information provided by respondents about their experienced uncertainty and interpreting and modelling this information as an integral part of the welfare estimation procedure is considered a promising alternative payment certainty calibration procedure for future welfare estimation. More research is needed in different public environmental domains to find further empirical support for this approach.

Acknowledgements

This study was conducted as part of the European Interreg IIIB NWE project Scaldit and funded by Rijkswaterstaat Zeeland. The survey work in France and Belgium was carried out by Arnaud Courtecuisse (French Water Agency Artois-Picardie), Ann Beckers and Lien van den Driessche (Flemish Environment Agency VMM), and Sandrine Dutrieux (Brussels Environment IBGE-BIM). A word of thanks goes to the International Scheldt Commission for its administrative support, Thijs Dekker, Marije Schaafsma and two anonymous reviewers for their useful comments on this paper. The usual disclaimer applies.

Annex 1: WTP question and certainty follow-up questions

Q. Are you as a household willing to pay every year € *X* in extra income tax over the next 10 years in order to reach a good ecological water quality status in 2015 in your part of the river basin?

Note: this money will only be used to finance the additional measures needed to reach a good ecological water quality status in 2015 in your part of the river basin as indicated in the map.

1 Yes
2 No

Q. Can you explain why you are willing to pay this specific amount of money, or if you are not willing to pay this specific amount of money why not?

……………………………………………………………………………………………………

……………………………………………………………………………………………………

Q. How certain are you that you will actually pay this specific amount of money, or if you are not willing to pay this specific amount of money how certain are you that you will not pay?

(please circle the appropriate percentage)

0% certain 10% 20% 30% 40% 50% 60% 70% 80% 90% 100% certain

Q. If you are not completely (100%) certain, can you explain why not?

……………………………………………………………………………………………………

……………………………………………………………………………………………………

Annex 2: The random effects probit model[13]

The probit model is represented as a probabilistic choice model $P_{ij}=Prob(V_{ij} + \varepsilon_{ij} > B)$ consisting of observable explanatory factors V and a random component ε due to unobservable factors. Individual i will accept to pay the bid price B if utility associated with the environmental change is higher than the price that has to be paid. The standard indirect utility function $U_{ij} = V_{ij} + \varepsilon_{ij}$ can be rewritten as $U_{ij} = \beta_j X_{ij} + \varepsilon_{ij}$ where the measurable component of utility is measured through a vector of *j* utility coefficients β associated with a vector of individual characteristics X_{ij}, and ε_{ij} captures the unobserved influences on an individual's choice. β is normally distributed in the population with mean b and covariance Ω: $\beta_j \sim N(b, \Omega)$. The random effects probit model assumes that model coefficients vary randomly across individuals instead of being fixed. β_j is in that case decomposed into its mean and deviations from its mean: $U_{ij} = bX_{ij} + \tilde{\beta}_j X_{ij} + \varepsilon_{ij}$ where $\tilde{\beta}_j = b - \beta_j$. The last two terms in the utility function are random. Denoting $\tilde{\beta}_j X_{ij} + \varepsilon_{ij} = \eta_{ij}$, utility becomes $U_{ij} = bX_{ij} + \eta_{ij}$. The covariance of η_{ij} now depends on both X_{ij} and Ω, such that the covariance differs over individuals.

[13] Based on Hausman and Wise (1978) and Train (2003).

References

Ajzen, I., Brown, T.C. and Carvajal, F. (2004). Explaining the discrepancy between intentions and actions: the case of hypothetical bias in contingent valuation. Personality and Social Psychology Bulletin, 30(9): 1108-1121.

Akter, S., Brouwer, R., Brander, L. and van Beukering, P. (2009). Respondent uncertainty in a contingent market for carbon offsets. Ecological Economics, 68: 1858–1863.

Alberini, A., Boyle, K., and Welsh, M. (2003). Analysis of contingent valuation data with multiple bids and response options allowing respondents to express uncertainty. Journal of Environmental Economics and Management, 45: 40-62.

Ariely, D., Loewenstein, G. & Prelec, D. (2003). "Coherent arbitrariness": Stable demand curves without stable preferences. Quarterly Journal of Economics, 118(1), 73-105.

Arrow, K., Solow, R., Portney, P.R., Leamer, E.E., Radner, R. and Schuman, H. (1993). Report of the NOAA Panel on Contingent Valuation. Federal Register, January 15, Vol.58, no.10: 4601-4614.

Ben-Akiva, M.E. and Swait, J.D. (1986). The akaike likelihood ratio index. Transportation Science, 20: 133-136.

Blumenschein, K., Blomquist, G.C., Johannesson, M., Horn, N. and Freeman, P. (2008). Eliciting willingness to pay without bias: Evidence from a field experiment. The Economic Journal, 118: 114-137.

Cameron, T.A. and James, M.D. (1987). Efficient estimation methods for closed-ended contingent valuation survey data. Review of Economics and Statistics, 69: 269-276.

Champ, P.A., Bishop, R.C., Brown,T.C. and McCollum, D.W. (1997). Using donation mechanisms to value nonuse benefits from public goods. Journal of Environmental Economics and Management, 33: 151-162.

Champ, P.A. and Bishop, R.C. (2001). Donation payment mechanisms and contingent valuation: An empirical study of hypothetical bias. Environmental and Resource Economics, 19: 383-402.

Chang, J-I., Yoo, S-H., and Kwak, S-J. (2007). An investigation of preference uncertainty in the contingent valuation study. Applied Economics Letters, 14: 691-695.

Greene, W.A. (2003). Econometric analysis. Prentice Hall, New Jersey.

Hausman, J.A. and Wise, D.A. (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogenous preferences. Econometrica, 46(2): 403-426.

Hoeffler, S. and Ariely, D. (1999). Constructing stable preferences: A look into dimensions of experience and their impact on preference stability. Journal of Consumer Psychology, 8(2), 113-139.

Johannesson, M., Liljas, B. and Johansson, P-O. (1998). An experimental comparison of dichotomous choice contingent valuation questions and real purchase decisions. Applied Economics, 30: 643-647.

Li, C-Z and Mattsson, L. (1995). Discrete choice under preference uncertainty: an improved structural model for contingent valuation. Journal of Environmental Economics and Management, 28: 256-269.

List, J.A. and Gallet, C.A. (2001). What experimental protocol influence disparities between actual and hypothetical stated values? Environmental and Resource Economics, 20: 241-254.

List, J.A. (2003). Does market experience eliminate market anomalies? Quarterly Journal of Economics, 41–72.

Loomis, J. and Ekstrand, E. (1998). Alternative approaches for incorporating respondent uncertainty when estimating willingness to pay: the case of the Mexican spotted owl. Ecological Economics, 27: 29-41.

Meyerhoff, J. and Liebe, U. (2006). Protest beliefs in contingent valuation: Explaining their motivation. Ecological Economics, 57: 583-594.

Ready, R.C., Whitehead, J.C. and Blomquist, G.C. (1995). Contingent valuation when respondents are ambivalent. Journal of Environmental Economics and Management, 29: 181-196.

Samnaliev, M., Stevens, T.H. and More, T. (2006). A comparison of alternative certainty calibration techniques in contingent valuation. Ecological Economics, 57: 507-519.

Sample, J. and Warland, R. (1973). Attitude and prediction of behavior. Social Forces, 51(3): 292-304.

Scaldit (2006). Scaldit report. Transnational analysis of the state of the aquatic environment of the international Scheldt river basin district: pilot project for testing the European guidance documents. Vlaamse Milieumaatschappij, Erembodegmem, Belgium. http://www.scaldit.org.

Shaik, S.L., Sun, L., and van Kooten, G.C. (2007). Treating respondent uncertainty in contingent valuation: A comparison of empirical treatments. Ecological Economics, 62: 115-125.

Train, K. (2003). Discrete choice methods with simulation. Cambridge University Press.

Veisten, K. and Navrud, S. (2006). Contingent valuation and actual payment for voluntarily provided passive-use values: Assessing the effect of an induced truth-telling mechanism and elicitation formats. Applied Economics, 38: 735-756.

Vossler, C.A., Ethier, R.G., Poe, G.L., and Welsh, M.P. (2003). Payment certainty in discrete choice valuation responses: Results from a field validity test. Southern Economic Journal, 69(4): 886-902.

Wang, H. (1997). Treatment of don’t know responses in contingent valuation surveys: a random valuation model. Journal of Environmental Economics and Management, 32(2): 219-232.

Welsh, M.P. and Poe, G.L. (1998). Elicitation effects in contingent valuation: comparisons to a multiple bounded discrete choice approach. Journal of Environmental Economics and Management, 36: 170-185.

Table 1: WTP - probit regression results

		Fixed Effects				Random Effects	
		Model 1		Model 2		Model 3	
Factor	**Value range**	$\hat{\beta}$	St err	$\hat{\beta}$	St err	$\hat{\beta}$	St err
Constant	-	-7.742	0.893	-7.534	0.862	-6.616	0.944
Belgium	0-1	-0.352	0.114	0.115[ns]	0.106	-0.592	0.145
Netherlands	0-1	-0.253[a]	0.125	0.196[b]	0.115	-0.491	0.154
Bid price	€5-250	-0.005	0.001	-0.007	0.001	-0.008	0.001
Age	18-92 years	0.011	0.003	0.007	0.003	0.013	0.003
Household income	€9-54*10^3	0.641	0.085	0.723	0.082	0.628	0.090
Importance	0-1	0.199[a]	0.084	0.181[a]	0.081	0.250	0.087
Familiarity	0-1	-0.326	0.094	-0.384	0.089	-0.315	0.101
Credibility (somewhat)	0-1	0.582	0.089	0.592	0.086	0.549	0.092
Credibility (completely)	0-1	0.714	0.149	0.803	0.146	0.655	0.155
Certainty	0-100%	0.020	0.002	0.564[*]	0.117		
Certainty*bid price	€0-250	-0.004	0.002	-0.002[a]	0.001		
$\sigma_{certainty}$						0.714	0.134
$\rho_{certainty}$						0.338	0.084
Log likelihood		-641.043		-698.957		-625.466	
Wald chi-square		434.35		369.04		314.99	
McFadden R-square		0.318		0.257		0.335	
N		1378		1378		1378	

[a] $p < 0.05$; [b] $p < 0.10$; [ns] not significant ($p > 0.10$); [*] dummy where 1 is 100% certain, and 0 is less than 100% certain

Table 2: Mean WTP with certainty corrections

	Fixed Effects Probit							RE Probit
	Model 1					Model 2		Model 3
Summary statistics	Uncorrected	Certainty ≥50%	Certainty ≥60%	Certainty ≥70%	Certainty ≥80%	100% certain	<100% certain	
Mean WTP (€/year)	107.4	92.8	63.1	50.9	22.3	127.4	86.3	87.5
95% confidence interval	97.6-117.2	83.1-102.6	52.9-73.3	40.5-61.2	10.1-34.6	113.3-141.6	71.8-100.8	74.7-100.3
MSE[1]	0.201	0.202	0.211	0.218	0.238	0.168	0.218	0.662
N	1662	1662	1662	1662	1662	701	961	1662

[1] Mean squared error: $\frac{1}{n}\left(\sum_{i=1}^{n}(W\hat{T}P_i - WTP_i)^2\right)$ where $W\hat{T}P$ is predicted WTP and WTP observed WTP

Table 3: Main sources of payment uncertainty

	Share (%)
Uncertainty related to the good self and its future supply	
Feasibility reaching good ecological water status	7.2
Insufficient information about the good and its supply	2.3
Doubt effectiveness of measures to be taken	1.3
Subtotal	*10.8*
Future demand uncertainty	
Future household income	17.3
Rising other household expenditures in future	2.4
Future situation in general	1.8
Subtotal	*21.5*
Price uncertainty	
High cost price	13.6
Future development cost price/tax	2.0
Calculation cost price	1.8
Subtotal	*17.4*
Survey instrument uncertainty	
General survey instrument uncertainty	
Doubt own contribution to the problem	3.3
Appropriateness of the tax instrument	2.8
Doubt influence on political decision	2.1
Existence other possible solutions	1.6
Doubt whether households are the right target group	0.9
Doubt whether paying extra is the solution to this problem	0.4
Subtotal	*11.1*
Policy scenario uncertainty	
Disbelief that the money will be spent on water quality improvements	13.0
Mistrust of the government	4.9
Control over how money will be spent and monitoring results	1.0
Subtotal	*18.9*
Uncertainty market conditions	
Whether polluters will pay	9.5
Whether everybody else will pay too	3.3
What surrounding countries will do	2.1
Subtotal	*14.9*
Other reasons	*5.4*
Total	*100.0*

Table 4: Payment certainty - ordered probit regression results

Factor	Value range	Pooled model $\hat{\beta}$	Pooled model St err	Yes responses only $\hat{\beta}$	Yes responses only St err	No responses only $\hat{\beta}$	No responses only St err
Imperfect knowledge/information							
Familiarity	0-1	-0.371	0.072	-0.355	0.108	-0.351	0.108
Supply related							
Credibility (completely)	0-1	0.291	0.101	0.224[b]	0.128	0.162[ns]	0.194
Demand related							
Household income	€9-54*10^3	0.500	0.065	0.532	0.096	0.172[b]	0.101
Price related							
Bid price	€5-250	-0.011	0.001	-0.008	0.002	-0.001[ns]	0.002
Bid price squared	€25-62500	0.337*10^{-4}	0.569*10^{-5}	0.243*10^{-4}	0.882*10^{-5}	0.955*10^{-5} [ns]	0.881*10^{-5}
Survey instrument related							
Protest	0-1	0.978	0.069	1.449	0.095	0.793	0.118
Other respondent characteristics							
Gender	0-1	-0.136[a]	0.067	-0.269	0.093	-0.023[ns]	0.104
Age	18-92 years	-0.006	0.002	-0.008	0.003	-0.010	0.003
Belgium	0-1	1.082	0.083	0.474	0.114	2.181	0.164
Netherlands	0-1	0.952	0.090	0.248[a]	0.122	2.149	0.176
Log likelihood		-1586.432		-749.430		-655.920	
Likelihood ratio chi-square		546.46		355.78		310.48	
McFadden R-square		0.147		0.192		0.191	
N		1434		812		622	

[a] $p < 0.05$; [b] $p < 0.10$; [ns] not significant ($p > 0.10$)

Figure 1: Probability distribution function yes votes and payment certainty yes and no votes